# Labour Market Risks

Borooah, Vani

July 2019

# Chapter 2
# Labour Market Risks

## 2.1 Introduction

Every time a job-seeker applies for a job he/she runs the risk of not getting it. However, these risks may not be uniformly distributed across job-seekers: some have a better chance of jumping the hurdles that serve as obstacles to employment; others have a higher chance of stumbling. The important question to ask relates to the determinants of such risk. In particular, does this risk differ significantly between job-seekers from different groups: gender, religion, or caste?

This chapter uses a famous result in statistics, Bayes' Theorem, to make explicit the concept of risk and to explain why, under this theorem, different groups might have different rates of success in securing employment. The theoretical results are buttressed by data from India relating to two periods: 2011–12 and 1999–2000. These data are used, in subsequent sections, to quantify the concept of risk set out in the earlier part of the chapter.

In addition to differences in risk faced by job-seekers from different groups in the labour market, there is also the question of overall inequality — embracing the experiences of all the groups — in the labour market. How is such aggregation to be arrived at? The last part of this chapter sets out an ingenious method of aggregation — with credit due to Theil (1967) and Bourguignon (1979) — involving nothing more than the arithmetic and geometric mean of group success rates in securing jobs. Applying this aggregation to Indian data reveals the egregiously high levels of inequality associated with both desirable (regular salaried) and undesirable (casual wage) jobs.

## 2.2 Bayes' Theorem and Dominant and Subordinate Groups

The Reverend Thomas Bayes, an 18th century Presbyterian minister, proved what, arguably, is the most important theorem in statistics.[1] Bayes' Theorem states that the probability of a hypothesis being true (event *T*), *given that the data have been observed* (event *A*), is the probability of the hypothesis

[1] See *The Economist*, "In Praise of Bayes", 28 September 2000, https://www.economist.com/science-and-technology/2000/09/28/in-praise-of-bayes

being true, *before any data have been observed*, times an "updating factor". The theorem is encapsulated by the well-known equation:

$$P(T \mid A) = \frac{P(A \mid T)}{P(A)} \times P(T) \qquad (2.1)$$

where: $P(T)$ represents the *prior* belief that the hypothesis is true *before the data have been observed*; $P(A)$ is the probability of observing the data, *regardless of whether the hypothesis is true or not*; $P(A|T)$ is the probability of observing the data, given that the hypothesis is true, and $P(A \mid T) / P(A)$ is the Bayesian "updating factor" which translates one's *prior* (that is, *before* observing the data) belief about the hypothesis's validity into a *posterior* (that is, *after* observing the data) belief .[2]

Bayes' theorem has been extensively applied in law and in medicine. For example, in the area of law it has shed light on the so-called "prosecutor's fallacy" whereby a prosecutor argues that since the probability of observing a particular piece of evidence (say, blood type identical to that found at the scene of the crime), *under the assumed innocence of the defendant*, is very small (that is, $P(A \mid T)$ is low), the probability of the defendant being innocent, *given that his blood type matches that at the crime scene*, must also be very small (that is, $P(T \mid A)$ must also be low). This fallacious reasoning stems, of course, from assuming that the ratio $P(A \mid T) / P(A)$ in equation (2.1) is equal to unity (Thompson and Schumann, 1987; Aitken, 1996).

In medicine the theorem has, for example, been used to analyse the efficacy of breast screening. Proponents of screening would argue, on the basis of the "screening fallacy", that because the probability of the screen returning a positive result, *given that the patient has cancer*, is large (that is, $P(A \mid T)$ is high), the probability of the patient having cancer, *given that the screen returns a positive result*, must also be large (that is, $P(T \mid A)$ must also be high). This fallacious reasoning stems, of course, from assuming that the ratio $P(T) / P(A)$ in equation (2.1) is equal to unity. However, if the proportion of persons with cancer in the population, relative to the proportion of positive screen results, is small (i.e. $P(T) / P(A)$ in equation (2.1) is low) then $P(T \mid A)$ could be

[2] The updating factor is the ratio of the probability of observing the data when the theory is true, to that of observing the data regardless of whether the theory is true or false: $P(A) = P(A \mid T)P(T) + P(A \mid \bar{T})P(\bar{T})$, $\bar{T}$ being the event that the theory is false.

appreciably smaller than $P(A|T)$. The size of this difference represents cancer "over diagnosis" and has recently been estimated at 10% (Zackrisson *et al.*, 2006). In effect, 1 in 10 women diagnosed with breast cancer would not require treatment.

These ideas can also be applied to the labour market. Suppose that a number of candidates, vying for jobs with a particular employer, have an innate suitability for employment (call it "employability"). Some of them are "employable" (event *T)* while others are "unemployable" (event $\tilde{T}$ ). Neither of these qualities, "employable" or "unemployable", is directly observable. Instead, candidates are admitted on the basis of a selection procedure such that candidates who pass the tests embodied in the procedure (event *A*) are offered jobs while and those who fail (event $\tilde{A}$ ) are rejected.

The tests are so designed that the probability (or, equivalently, the likelihood) of a candidate passing the test, *if he/she is "employable"*, is $\alpha$, and denoted $P(A|T)=\alpha$ .[3] The *sensitivity* of the tests is defined by $\alpha$ : the more sensitive the test (that is, the larger the value of $\alpha$), the greater the likelihood that an "employable" candidate will pass the test. Similarly, the *specificity* of the tests, $\beta$, is defined as the likelihood that an "unemployable" candidate will fail the test: $P(\tilde{A}|\tilde{T})=\beta$ .

Following from this, *1-specificity = 1-*$\beta$ = $1-P(\tilde{A}|\tilde{T})=P(A|\tilde{T})$ is the probability of a *false positive*: the probability that a candidate who is "unemployable" will pass the test. [4] Similarly, *1-sensitivity = 1-*$\alpha$ = $1-P(A|T)=P(\tilde{A}|T)$ is the probability of a *true negative*: the probability that an "employable" candidate will fail the test.[5] These four possibilities are set out in Table 2.1.

<Table 2.1>

However, the question of interest is not what is the likelihood (probability) of an "employable" candidate passing the test ( $P(A|T)$ ) but, rather, what is the likelihood that a candidate who passes the test will be "employable" ( $P(T\,|\,A)$ ).[6] It is important to emphasise that these are two

[3] *P* in this term represents probability and the symbol | denotes that the event following | has already occurred.
[4] "False positive": an unemployable candidate (false) passes the test (positive).
[5] "True negative": an "employable" candidate (true) fails the test (negative).
[6] Similarly, the question of interest is not what is the likelihood of an "unemployable" candidate failing the test, $P(\tilde{A}|\tilde{T})$, but, rather, what is the likelihood that a candidate who fails the test will be "unemployable", $P(\tilde{T}\,|\,\tilde{A})$.

separate and distinct questions and the strength of Bayes' Theorem is that it is able to provide an answer to the second question by linking it to the first.

As discussed above, Bayes' Theorem states that the probability of a theory being true (event *T*: a candidate is "employable"), *given that the data have been observed* (event *A*: he/she passed the selection test) is given by equation (2.1), above. After the data have been observed, the Bayesian "updating factor" in equation (2.1), $P(A|T)/P(A)$, translates one's *prior* belief about the theory's validity into a *posterior* belief or, in the context of the labour market example, translates one's *prior* belief about a candidate's employability into a *posterior* belief.

The probability of passing the test (event *A*) is the weighted sum of the probabilities of a "true positive" (an "employable" candidate [true] passes the test [positive]) and a "false positive" (an "unemployable" candidate [false] passes the test [positive]), the weights being the strength of one's prior belief, $P(T)$, where $P(\tilde{T}) = 1 - P(T)$:

$$P(A) = P(A \cap T) + P(A \cap \tilde{T}) = P(T) \times \overbrace{P(A|T)}^{\text{prob of true positive}} + P(\tilde{T}) \times \overbrace{P(A|\tilde{T})}^{\text{prob of false positive}} \tag{2.2}$$

where: $\tilde{T}$ is the event that the candidate was "unemployable". Substituting the expression in (2.2) into equation (2.1) yields:

$$P(T \mid A) = \frac{P(T) \times P(A|T)}{P(T) \times P(A|T) + P(\tilde{T}) \times P(A|\tilde{T})} \tag{2.3}$$

Suppose that, prior to the selection test being administered, there is a prior belief that one in 10 candidates will be "employable", that is, $P(T) = 0.1$. Suppose also that the test is such that its sensitivity, $\alpha = P(A|T) = 0.95$ and its specificity, $\beta = P(\tilde{A}|\tilde{T}) = 0.85$ implying that the probability of a false positive is $1 - \beta = P(A|\tilde{T}) = 0.15$. This means that, on average, 95 out of 100 "employable" candidates, and 15 out of 100 "unemployable" candidates, will pass the test or, in other words, the probabilities of a "true positive", $P(A|T)$, and a "false positive", $P(A|\tilde{T})$, are, respectively, 95% and 15%.

Substituting these assumed values into equation (2.3) yields:

$$P(T \mid A) = \frac{0.1 \times 0.95}{0.1 \times 0.95 + 0.9 \times 0.15} = 0.413 \tag{2.4}$$

or, in other words, there is a 41% chance that a candidate passing the test will be 'employable' suggesting that the *selection fallacy*, which arose from confusing $P(A|T)[=0.95]$ with $P(T \mid A)[=0.41]$, is not negligible.

The value of $P(T \mid A)$ will be smaller for a lower, and larger for a higher, *prior belief* about the likelihood of the candidates being employable: *P(T)*. For example, $P(T \mid A) = 0.25$ when $P(T) = 0.05$ and $P(T \mid A) = 0.613$ when $P(T) = 0.2$. More generally, $\partial P(T \mid A)/\partial P(T) < 0$.

If there is to be no *selection fallacy*, meaning that $P(T \mid A) = P(A|T)$ — or, in words, the probability of being "employable" (*T*), if one passes the test *(A)*, is equal to the probability of passing the test (*A*), if one is "employable" *(T)* — then $P(T) = 0.75$: the prior belief is that three out of four candidates will be "employable".[7] If $P(T) < 0.75,$ then there will be a selection fallacy meaning that: $P(T \mid A) < P(A|T)$ or, the probability of being employable (*T*), if one passes the test, is less than the probability of passing the test (*A*), if one is employable.

### 2.3 Employability and Social Groups

The forgoing analysis assumed that the specificity of the test, the likelihood that an unemployable person would fail the test, $P(\tilde{A}|\tilde{T}) = 0.85$, was exogenously given and could not be affected by the persons either taking or conducing the test. A corollary of this is that the probability of a false positive, the likelihood that an unemployable person would pass the test, $P = (A|\tilde{T}) = 1 - P(\tilde{A}|\tilde{T}) = 0.15$, would also be outside the control of the candidates and the testers. In the case of selection tests, however, *discrimination* between candidates from different group can result in inter-group differences in the likelihood of a false positive or, in other words, differences between "unemployable" candidates from different groups in their likelihood of passing the test.

[7] $P(T \mid A) = P(A|T) = 0.95 \Rightarrow 0.95 = \dfrac{P(T) \times 0.95}{P(T) \times 0.95 + (1 - P(T)) \times 0.15} \Rightarrow P(T) = \dfrac{0.15}{0.2} = 0.75$.

Suppose that persons in the labour market belong to one or the other of two groups — a socially *dominant* group and a socially *subordinate* group — and suppose that 75% of job-seekers are from the dominant group and 25% are from the subordinate group. Suppose that the assumed 85% specificity of the test could, *with discrimination in favour of the dominant group*, be different for job-seekers from the dominant and subordinate groups. Then under such a discriminatory policy, the overall specificity, $\beta = P(\tilde{A} \mid \tilde{T})$ is:

$$\beta = P(\tilde{A} \mid \tilde{T}) = 0.75 \times P(\tilde{A}_D \mid \tilde{T}_D) + 0.25 \times P(\tilde{A}_S \mid \tilde{T}_S) \qquad (2.5)$$

In equation (2.5), $\tilde{A}_D$ and $\tilde{A}_S$ are, respectively, the events that job-seekers from the dominant and subordinate groups fail the selection tests and $\tilde{T}_D$ and $\tilde{T}_S$ are, respectively, the events that job-seekers from the dominant and subordinate groups are "unemployable".

Suppose that employers wanted to keep the overall specificity constant at 85%, $P(\tilde{A} \mid \tilde{T}) = 0.85$, but, by discriminating in favour of the dominant group, and *ipso facto* against the subordinate group, to reduce the specificity for job-seekers from the dominant group to 82% and thereby to raise the probability of a false positive, $P(A_D \mid \tilde{T}_D)$, from 15% to 18%.

Then, from equation (2.5), in order to keep the overall specificity constant at $P(\tilde{A} \mid \tilde{T}) = 0.85$, the specificity of the selection test for job-seekers from the subordinate group would have to rise to 94%.[8] This means that the probability of a false positive, $1 - P(\tilde{A} \mid \tilde{T}) = P(A \mid \tilde{T})$, *which is the probability that an unemployable job-seeker would receive a job offer*, would, under discrimination, be 18% for job-seekers from the dominant group but only 6% for job-seekers from the subordinate group. If the specificity for job-seekers from the dominant group was reduced to 80%, then the probabilities of a false positive would be 20% and 0% for job-seekers from, respectively, the dominant and subordinate groups.[9]

---

[8] $P(\tilde{A}_S \mid \tilde{T}_S) = \dfrac{0.85 - 0.75 \times 0.82}{0.25} = 0.94$.

[9] $P(\tilde{A}_S \mid \tilde{T}_S) = \dfrac{0.85 - 0.75 \times 0.80}{0.25} = 1$

### 2.3.1 *Inter-Group Disparities in Employment Outcomes*

Employers could distinguish between job-seekers from different groups for two main reasons: they may have an innate dislike for people from the subordinate group (Becker, 1993) and/or they may feel, rightly or wrongly, that, on average, they are not such good workers as those from the dominant group (Arrow, 1972a, 1972b, 1973). Here the focus is on the second reason as the basis for treating job-seekers from different groups differently. This inter-group difference in the treatment of job-seekers will, as argued in some detail below, lead to inter-group disparities in employment outcomes.

In the analysis so far, employers' prior belief was that there was no difference in the ability between job-seekers from the two groups. However, it is possible that employers believe, rightly or wrongly, that job-seekers from the subordinate group are, on average, less employable than those from the dominant group. Consequently, their prior belief about the employability of job-seekers will be lower for the subordinate than for the dominant group of job-seekers: $P(T^S) < P(T^D)$, where these probabilities represent the prior belief that job-seekers from, respectively, the subordinate and dominant groups are "employable".

Suppose that the prior belief of employers is that there is 5% chance of job-seekers from the subordinate group, but a 20% chance of job-seekers from the dominant group, being employable: that is, $P(T^S) = 0.05$ and $P(T^D) = 0.2$. Given this prior belief, equation (2.3) computes that, if job-seekers passed the selection tests, there would be a 61% chance of them being employable if they belonged to the dominant group but only a 25% chance of them being employable if they belonged to the subordinate group: $P(T_D \mid A_D) = 0.61$ and $P(T_S \mid A_S) = 0.25$ .

These differences in employability between successful job-seekers from the two groups, stemming from differences in employers' prior beliefs, then offer employers an incentive to make it relatively difficult for the subordinate group's job-seekers, and relatively easy for the dominant group's job-seekers, to find employment. They achieve this by skewing the selection tests such that the specificity for the dominant group is less than that for the subordinate group, $P(\tilde{A}_D \mid \tilde{T}_D) < P(\tilde{A}_S \mid \tilde{T}_S)$, or, equivalently, that the probability of a false positive is larger for job-seekers from the dominant group than for those from the subordinate group: $P(A_D \mid \tilde{T}_D) > P(A_S \mid \tilde{T}_S)$.

At the extreme, when $P(\tilde{A}_D \mid \tilde{T}_D) = 0.8$ with overall specificity $P(\tilde{A} \mid \tilde{T}) = 0.85$, the probability of a false positive for job-seekers from the subordinate group is reduced to zero: it would be impossible for an "unemployable" job-seeker from this group to find employment but it would be relatively easy (20%) for an "unemployable" job-seeker from the dominant group to do so.

**2.3.2** ***A Numerical Example***

Now suppose that of 1,000 job-seekers, 750 are from the dominant group and 250 from the subordinate group. The nature of the selection process is that an "employable" candidate will have a 95% of passing the selection tests (and, therefore, a 5% chance of failing to land a job) and an "unemployable" candidate will have an 85% chance of failing the test (and, therefore, a 15% chance of getting a job). The prior belief is that there is no difference in ability between job-seekers from the two groups, so that if there were 100 vacancies, 75 and 25 of these would go to, respectively, the dominant and subordinate group candidates.

Suppose now that employers — acting on the fact that their prior belief about the employability of job-seekers is lower for those from the subordinate group than those from the dominant group — skew the selection process so that the probability of a false positive is 6% for the subordinate group's, and 18% for the dominant group's, job-seekers, under the umbrella of an *overall* 15% probability of a false positive. Then we can make the following calculations.

1. Of the 750 dominant group's job-seekers, the prior belief, $P(T_D) = 0.2$, is that 150 are "employable" and 600 are "unemployable". Of the 150 "employable" and the 600 "unemployable" job-seekers from the dominant group, respectively, 142 (95% of 150) and 108 (18% of 600) candidates will pass the test. This means that, of the 250 dominant group's job-seekers who pass the test, 57% will be "employable" and 43% will be "unemployable" candidates.
2. Of the 250 backward class candidates, the prior belief is that, $P(T_S) = 0.05$, 12 are "employable" and 238 are "unemployable". Of the 12 "employable", and the 238 "unemployable", job-seekers from the subordinate group, respectively, 11 (95% of 12) and

14 (6% of 238) candidates will pass the test. This means that of the 25 subordinate group's job-seekers who pass the test, 44% will be "employable" and 56% will be "unemployable".

Because of differences in employers' prior beliefs about the employability of job-seekers from the dominant and subordinate groups, the ratio of "unemployable" job-seekers that passed the test from the dominant group, to those from the subordinate group, is nearly 8:1 (108:14). Furthermore, as a consequence of the difference between the two groups in their capacity to generate false positives, the proportion of successful job-seekers from the dominant and subordinate groups are, respectively, 33% (250/750) and 10% (25/250) and this does not reflect their respective shares of 75% and 25% in the job-seekers' population.

It is important to stress that this disparity between the shares of job-seekers from the subordinate and dominant groups in employment, respectively, 10% and 33%, and their shares in the population of job-seekers, respectively, 25% and 75%, stems entirely from differences in employers' prior beliefs about the employability of job-seekers from the two groups. *But disparity does not necessarily stem from discrimination.*

Employers may hold these disparate prior beliefs either rightly or wrongly. If, indeed, there are genuine differences between job-seekers from the two groups in their endowment of employment-friendly characteristics, then inter-group differences in employment outcomes represent the "unequal treatment of unequals" and can be justified on grounds of "business necessity": there is disparity but there is no discrimination. On the other hand, if differences between job-seekers from the two groups in their endowment of employment-friendly characteristics are but a figment of employers' imagination, then inter-group differences in employment outcomes represent the "unequal treatment of equals" and deserve to be labelled, indeed deplored, as "discrimination".

### 2.4 Defining Labour Market Risk using Bayes' Theorem

The preceding section argued that because of disparities between job-seekers from the dominant and subordinate groups in their treatment by employers — specifically in terms of prior beliefs about their "employability" — subordinate group job-seekers faced a higher risk of labour market failure than their counterparts from the dominant group. As the numerical example of the previous section showed, there was considerable underrepresentation of successful job-seekers from the subordinate group matched by considerable overrepresentation of successful job-seekers from the dominant group. This section explores in greater detail the concept of the risk of labour market failure or, more succinctly, *labour market risk*.

One definition of the labour market risk associated with belonging to a particular group (*X*) is the ratio of the likelihood that the job-seeker from that group gets the job (event *A*), to the likelihood that he/she does not get it (event $\tilde{A}$), *given that he/she belongs to group X*. This ratio is, hereafter, referred to as the *risk ratio* (RR) and is denoted by $\rho$, where:

$$\rho = \frac{P(A|X)}{P(\tilde{A}|X)} = \frac{P(X|A)}{P(X|\tilde{A})} \times \frac{P(A)}{P(\tilde{A})} = \frac{P(X|A)}{P(X|\tilde{A})} \times \frac{P(A)}{1-P(\tilde{A})} = \Phi \frac{P(A)}{1-P(\tilde{A})} \qquad (2.6)$$

where: $\Phi = \dfrac{P(X|A)}{P(X|\tilde{A})} = \dfrac{\rho}{\lambda}$, where $\lambda = \dfrac{P(A)}{1-P(A)}$ is the *odds ratio* (OR) that is, the ratio of the likelihood of getting, to not getting, a job, regardless of group affiliation.

The term $\Phi$ in equation (2.6) is the so-called *Bayes Factor* (BF) applied to job-seekers from the subordinate group. The Bayes Factor is a measure of whether the data (the job-seeker is from group *X*) are more likely to be observed under one outcome (*A*: he/she gets the job) than under the alternative outcome ($\tilde{A}$: he/she does not get the job): $\Phi < 1 (> 1)$ signifies that the likelihood of a job-seeker belonging to group *X* is lower (higher) when he/she gets the job compared to when he/she does not get job. It tells us by how much we should alter our *prior belief* that a job-seeker will get a job with probability $P(A)$, and fail to get it with probability $P(\tilde{A}) = 1 - P(A)$, *in the light of the data that the job-seeker is from group X.*[10]

### 2.4.1 *The Inverse Bayes' Factor*

The risk ratio, $\rho$ in equation (2.6), measured the odds of the null hypothesis being "true" (*A:* the job-seeker gets the job) to it being "false" ($\tilde{A}$: the job-seeker does not get the job) *under a particular set of data* which, in this case, is that the job-seeker was from group *X*. In this formulation of risk, the data applicable to the different outcomes (getting or not getting a job) are the same (the job-seeker was from group *X*). An alternative view of risk is obtained by posing the following question: given

[10] See Matthews (2000).

two rival scenarios — first, the job-seeker was from group *X* and second, he/she was from group *Y* — what is the ratio of the probabilities of getting a job in these different scenarios?

In this case, the risk ratio of belonging to group *X* is the ratio of the likelihood that a job-seeker got employment as a member of group *X*, to the likelihood that a job-seeker found employment as a member of group *Y*. Here the outcome is the same (*A:* a job-seeker gets the job) but the data that are input are different (group *X* or *Y*). In order to answer this question, the relevant risk ratio (represented by $\sigma$) is $\sigma = \frac{P(A|X)}{P(A|Y)}$. Hereafter, $\sigma$ is referred to as the *inverse risk ratio* (IRR): given two different "pieces" of information — a job-seeker is from group *X* or group *Y* — what is the ratio of the probabilities of getting a job?

In turn, one can expand $\sigma$ so that:

$$\sigma = \frac{P(A|X)}{P(A|Y)} = \frac{P(X|A)P(A)}{P(X)} \times \frac{P(Y)}{P(Y|A)P(A)} = \frac{P(X|A)}{P(Y|A)} \times \frac{P(Y)}{P(X)} = \Psi \frac{P(Y)}{P(X)} \qquad (2.7)$$

where: $\Psi = \frac{P(X|A)}{P(Y|A)} = \frac{\sigma}{\mu}$ where $\mu = \frac{P(Y)}{P(X)}$ is the *inverse population ratio* (IPR): the ratio of the number of job-seekers from group *Y* to that from group *X*.

The term $\Psi$ in equation (2.7) is the *Inverse Bayes Factor* (IBF) applied to the job-seeker that got the job. The IBF is the odds of the null hypothesis being true (the job was secured) under one set of data (the job-seeker was from group *X*), against it being true (the job was secured) under the obverse set of data (the job-seeker was from group *Y*). If $\Psi < 1 (> 1)$ then, given that the hypothesis is true (the job is secured), we are less (more) likely to observe one set of data (the job-seeker was from group *X*) than the complementary set of data (the job-seeker was from group *Y*).

## 2.5 Risk Ratio and Bayes' Factor Calculations for the Indian Labour Market

This section puts empirical flesh on the theoretical skeleton of the previous section using unit record data from the latest available round (68th round: July 2011–June 2012), and an earlier round (55th round: July 1999–June 2000), pertaining to a decade earlier, of the National Sample Survey (NSS) of Employment and Unemployment. The NSS are carried out by the National Sample Survey Organisation (NSSO), an autonomous agency under the Government of India's Ministry for Statistics.

As described in Tendulkar (2007), the NSS adopt a personal interview method with a schedule of items on which information is elicited from members of each sampled household. Obtaining this information relies on a retrospective recall of certain items, for example, expenditure undertaken over the previous year on education. These interviews are conducted in the context of a two-stage stratified random sample in which villages or urban blocks comprise the first stage units (FSU) and the households in the selected villages/units constitute the second stage units (SSU).[11] The NSS employment and unemployment data give the distribution of its respondents — who are distinguished by various characteristics, including their social group and educational levels — between different categories of economic status: (i) regular salaried or wage employees; (ii) casual wage employees; (iii) own-account workers (including employers); (iv) unpaid family workers; (v) domestic duties; (vi) unemployed, that is, actively seeking and available for work.[12]

Since a major purpose of this chapter is to study the representation of social groups in the various categories of employment status (i–vi, above), it was particularly important to construct meaningful social groups such that each person in the estimation sample was placed in one, and only one, of these groups. The NSS categorises persons by four "social" groups — Scheduled Tribes (ST); Scheduled Castes (SC); Other Backward Classes (OBC); and "Other" — and simultaneously by eight "religion" groups — Hindu; Islam; Christianity; Sikhism; Jainism; Buddhism; Zoroastrianism; "Other". The fact that Muslims, too, have their "backward classes"' and "forward" classes, with a conspicuous lack of inter-marriage between the two groups, means that it is sensible to separate Muslims into two groups: Muslims from the OBC and non-OBC Muslims (Trivedi *et al.*, 2016).[13]

Combining the NSS "social group" and "religion" categories, households were subdivided into the following groups which were then used as the basis for analysis in this chapter:

1. Scheduled Tribes (ST): these comprised 8.5% of the households surveyed in the 68th round.

---

[11] Consequently, as Tendulkar (2007) observes, "the availability of well-trained and motivated staff poses the biggest constraint given the budget provided by the government, their procedures of recruitment, and their distribution in a multilingual country" (p. 368).

[12] Excluded from this chapter's analysis were pensioners, rentiers, and remittance receivers and also those who earned their living from begging or prostitution.

[13] The Sachar Committee (2006) refers to the caste system applying also to Muslims with the *ashraf* (meaning "noble") referring to high-born Muslims and converts to Islam from Hindus of the higher castes, and the *ajlaf* (meaning "degraded" or "unholy") referring to converts to Islam from the lower castes. So, among Muslims too there are "backward classes" (like *Teli*, *Ansari*, *Qureshi*, *Banjara*, *Nai* etc.) and forward classes.

2. Scheduled Castes (SC): these comprised 18.8% of the households surveyed in the 68th round with over 94% of these households being Hindu.
3. Other Backward Classes that were non-Muslim (OBC-NM): these comprised 37% of households in the 68th round with 96% of these households being Hindu.
4. Muslims: these comprised 13.8% of households in the 68th round.
5. Forward Castes (FC): these comprised 21.8% households in the 68th round with 92% of these households being Hindu, 4% being Sikh, 3% being Christian, and 1% being Jains.

Table 2.2 shows the representation in the six categories of employment status, enumerated above, for persons in the 68th and 55th rounds belonging to the five different social groups.[14] The information in the table relates to persons between the ages of 21 and 55 years and is separated by gender. The last column of Table 2.2 (under the column headed "Total") shows that, in the 68th round, 21.3% of males (aged 21–55) were regular salaried and wage employees (hereafter, RSWE), 28.6% were casual wage employees (CWE), 38.9% were own-account workers (OAW), 9.4% were unpaid family workers (UFW), and 1.6% were unemployed. By comparison to the 68th round, the 55th round shows that, a decade earlier, a smaller proportion of men (aged 21–55) were RSWE (19%), a higher proportion of men were CWE (30.6%), with a roughly similar proportions being OAW and UFW. For women aged 21–55, data from the 68th round showed that 61.6% were engaged in domestic duties (DD) with only 5.2% categorised as RSWE, 11.6% as CWE, 7.7% as OAW, 13% as UFW, and less than 1% as unemployed. By comparison to the 68th round, the 55th round showed that, a decade earlier, a lower proportion of women were RSWE (3.6%), OAW (5.1%) and UFW (11.7%), a higher proportion of women were CWE (15.5%), with roughly similar proportions engaged in DD.

<Table 2.2>

Underlying these aggregate trends, however, were differences in representation between the social groups in the different categories of employment status. In the 68th round, 32.4% of FC men — compared to 12% of men from the ST, 18.2% of men from the SC, 19.3% of men from the OBC-NM, and 17.5% of Muslim men — were RSWE. At the other extreme, 47.8% of SC men — compared to

[14] It is important, at the very outset, to draw attention to the fact that all the results reported in this book are obtained after grossing up the survey data using the observation-specific weights provided by the NSS for the 68th and 55th rounds.

12.3% of men from the FC, 27.3% of men from the OBC-NM, 28.8% of Muslim men, and 36.6% of men from the ST — were CWE. This pattern was unchanged from the 55th round except that, compared to the 68th round, a lower proportion of men from all the social groups were RSWE and a higher proportion were CWE. Similarly, data from the 68th round showed that 76% of Muslim women and 68.7% of FC women were engaged in DD, in contrast to 40.9% of ST women, 56.6% of SC women, and 59.5% of OBC-NM women; these figures were uniformly lower than those from the 55th round when 81.2% of Muslim women and 74.1% of FC women were in DD.

The first two panels of Table 2.3 show, for the 68th and 55throunds, the Bayes' Factor (BF) and risk ratio (RR) calculations, pertaining to equation (2.6), for three categories of employment status: RSWE, CWE, and OAW. These panels show that the odds ratio of being RSWE, CWE, and OAW were, respectively 0.27, 0.4, and 0.64 for the 68th round and 0.24, 0.44, and 0.61 for the 55th round. These odds ratios are the empirical counterpart of $\lambda$ in equation (2.6): they represent the odds of being in a particular status to not being in that status, irrespective of group membership. So, for example, in the 68th round, after grossing up, 52,970,668 persons were RSWE and 195,654,565 (=248,625,233-52,970,668) persons were not in RSWE yielding an odds ratio ($\lambda$ value) for RSWE of 0.27; similarly, in the 55th round, after grossing up, 47,207,270 persons were CWE and 107,260,330 (=154467600-47207270) persons were not in CWE yielding an odds ratio ($\lambda$ value) for CWE of 0.44.

<Table 2.3>

The numbers in the row labelled "Bayes' Factor" are the empirical counterpart of $\Phi$ in equation (2.6): this measures the ratio of the likelihood of a person belonging to a particular group, *if he/she attains a specific employment status*, to the likelihood of belonging to that group, *if he/she does not attain that employment status*. The estimated values of the BF for the 68th round show that the likelihood of attaining, to not attaining, RSWE status was as low as 0.14 for the ST and as high as 0.48 for the FC.[15] In contrast, for the 68th round, the likelihood of attaining, to not attaining, CWE status, was as high as 0.92 for the SC and as low as 0.14 for the FC. These figures echo those for the

[15] These figures were calculated as follows: after grossing up, there were 2,538,594 men from the ST in RSWE and 18,619,663 (=21158257-2538594) men from the ST *not* in RSWE yielding a BF=0.14 (=2,538,594/18,619,663). Similarly, after grossing up, there were 21,940,189 men from the SC in CWE and 23,976,299 (=45,916,488-21,940,189) men from the SC *not* in CWE yielding a BF=0.92 (=21,940,189 /23,976,299).

55th round: the BF for RSWE was highest for the FC and lowest for the ST, and the BF for CWE was highest for the SC and lowest for the FC.

As equation (2.6) shows, multiplying the BF ($\Phi$) by the odds ratio ($\lambda$), for a particular employment status, yields the risk ratio (RR) for that employment status. This is $\rho$ of equation (2.6). The RR represents the ratio of the likelihood of a person attaining a specific employment status, to not attaining it, *given that he/she is from a particular group*. Expressed as a percentage, the row labelled "Risk Ratio" in Table 2.3 shows that, for the 68th round, the likelihood of a person from the SC being RSWE was 6% of the likelihood of a person from the SC *not* being RSWE; by contrast, for the 68th round, the likelihood of a person from the FC being RSWE was 13% of the likelihood of a person from the FC *not* being RSWE. If one normalises on the RR of persons from the FC, then the row labelled "Normalised Risk Ratios" shows that the RR of an ST person being RSWE was only 28%, and the RR of an SC person being RSWE was 46%, of that of a FC person being RSWE. In contrast, for an "undesirable" status like CWE, Table 2.3 shows that, for the 68th round, the RR of a SC person being CWE was 36.7% and this was 6.6 times higher than the RR of a FC person being in CWE (5.6%).

The last two panels of Table 2.3 show, for the 68th and 55th rounds, the Inverse Bayes' Factor (IBF) and inverse risk ratio (IRR) calculations, pertaining to equation (2.7), for the three categories of employment status: RSWE, CWE, and OAW. The first line for each of these categories shows the values of the inverse population ratio (IPR). These are the counterpart of $\mu$ in equation (2.7): they represent the total number of persons in the sample from group *Y* — which is the comparator group of persons from the FC — relative to the number in group *X* (ST, SC, OBC-NM, and Muslims). So, for example, there were, after grossing up, a *total* of 57,855,811 persons from the FC and 45,916,488 persons from the SC in the 68th round yielding a IPR of 1.3 (=57,855,811/ 45,916,488) for the SC. Similarly, there were, after grossing up, a *total* of 44,450,440 persons from the FC and 50,006,943 persons from the OBC-NM in the 55th round yielding a IPR of 0.89 (=44,450,440 /50,006,94) for the OBC-NM.

The numbers in the row labelled "Inverse Bayes' Factor" are the empirical counterpart of $\psi$ in equation (2.7): this measures the ratio of the likelihood of a person belonging to a specific employment status, *if he/she is from one group (group X)*, to the likelihood of belonging to that status, *if he/she is from a different group (group Y)*. If the reference group (group *Y*) is taken to be the FC, then the estimated values of the IBF for the 68th round show that the likelihoods for ST, SC, OBC-NM, and Muslim men of being RSWE, were, respectively: 14%, 45%, 95%, and 29% of that of men from the FC.[16] In contrast, for the 68th round, the likelihoods for ST, SC, OBC-NM, and Muslim men of being CWE, were, respectively: 1.1, 3.1, 3.5, and 1.3 times that of the likelihood of men from the FC being CWE. An important feature of a comparison between the IBF figures from the 55th and 68th round is that, in the decade between the rounds, the IBF of men from all the groups (ST, SC, OBC-NM, Muslims) has increased relative to that of FC men or, in other words, *given that men are RSWE*, the likelihood of them being ST, SC, OBC-NM, or Muslim, relative to the likelihood of them being FC, was higher in the 68th round than in the 55th round. That is progress!

As equation (2.7) shows, multiplying the IBF ($\psi$) by the IPR ($\mu$), for a particular employment status, yields the inverse risk ratio (IRR) for that employment status. This is $\sigma$ of equation (2.7). The IRR represents the ratio of the likelihood that a person is in a specific employment status under two different scenarios: (i) he/she is from group X, or (ii) he/she is from group Y. Expressed as a percentage, the row labelled "Inverse Risk Ratio" in Table 2.3 shows that, for the 68th round, the likelihoods ST, SC, OBC-NM, and Muslim men being RSWE were, respectively, 37%, 56%, 60% and 54% of the likelihood of FC men being RSWE. In contrast, for an "undesirable" status like CWE, Table 2.3 shows that, for the 68th round, the likelihoods of ST, SC, OBC-NM, and Muslim men being CWE were, respectively, 3, 3.9, 2.2 and 2.3 times the likelihood of FC men being CWE. An important feature of a comparison between the IRR figures from the 55th and 68th round is that, in the decade between the rounds, the IRF of men from all the groups, but, in particular, SC and OBC-NM men, has increased relative to that of FC men or, in other words, the likelihood for SC and OBC-NM men,

[16] These figures were calculated as follows: after grossing up, there were 8,345,334 men from the SC and 18,747,396 men from the FC in RSWE yielding, for SC men, an IBF=0.45 (=8,345,334/18,747,396). Similarly, after grossing up, there were 5,461,887 Muslim men in RSWE yielding, for Muslim men, an IBF=0.29 (=5,461,887/18,747,396).

relative to the likelihood for FC men, of being in RSWE has increased between the 55th and 68th rounds. That, too, is progress!

### 2.6 Measuring Inequality in Access to Employment

The preceding sections examined inequality in outcomes across different types of status with respect to specific groups: ST, SC, OBC-NM, Muslims, and FC. This section examines the question of aggregation: how should the experiences of the different groups be aggregated so as to arrive at an overall measure of inequality with respect to specific employment outcomes?

The most usual concept of "unfair access" by a group to a particular "facility" is that there is disproportionality between its representation in the population and in the facility. However, when there are many groups, the relevant question is how to merge these group disproportionalities into a single measure of *access inequality*. Ideally such a measure should satisfy the "Pigou-Dalton condition" which, applied to the present study, requires that an increase in numbers of deprived persons, *at the expense of an equal reduction in the number of non-deprived persons*, would reduce access inequality.[17]

Suppose that a population of $N$ persons is divided into $M$ mutually exclusive and collectively exhaustive groups with $N_m$ ($m=1...M$) persons in each group such that $N_m$ and $H_m$ are the numbers in *each* group in, respectively, the overall population and in the population of persons who are employees of a particular type, say RSWE. Then $N = \sum_{m=1}^{M} N_m$ and $H = \sum_{m=1}^{M} H_m$ are, respectively, the total number of persons in the overall population and of the number of persons who are RSWE.

One way of measuring inequality in a variable is by the natural logarithm of the ratio of the arithmetic mean of the variable to its geometric mean.[18] As Bourguignon (1979) demonstrates, such a measure satisfies *inter alia* the Pigou-Dalton condition. This idea translates very naturally from its usual application to income inequality, to measuring the degree of inequality associated with employment outcomes in which people belonging to different population groups meet with different degrees of success of being RSWE.

[17] In the language of inequality analysis this transfer *from* an "access-rich" group *to* an "access-poor" group constitutes a progressive transfer and, by virtue of this, is inequality reducing.
[18] See Theil (1967), Bourguignon (1979) and Borooah (2001).

The variable of interest is the *access rate* to RSWE of persons from group $m$ — defined as the proportion of persons from that group who are in RSWE — and it is inequality in the distribution of this rate between the $M$ groups that is sought to be measured. This inequality is referred to as “access inequality”.

The success rate (SR) of group $m$ (denoted $e_m$) is $e_m = H_m / N_m,\ 0 \le e_m \le 1$. Then the arithmetic and geometric means of $e_m$ are, respectively:

$$\overline{e} = \sum_{m=1}^{M} e_m n_m \quad \text{and } \hat{e} = \prod_{m=1}^{M} (e_m)^{n_m} \ \ where\ n_m = N_m / N,\ \ \sum_{m=1}^{M} n_m = 1 \qquad (2.8)$$

so that the measure of access inequality is:

$$J = \log(\overline{e} / \hat{e}) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log(e_m) > 0 \qquad (2.9)$$

since, by the property of means, the arithmetic mean is greater than or equal to the geometric mean.

<Table 2.4 >

If one takes the five social groups used in this study (ST, SC, OBC-NM, Muslims, and FC) then Table 2.4 shows the SR of men from the different groups in three different types of employment: RSWE, CWE, and OAW. This success rate (defined by $e_m$, above) is the proportion of men from that group who are in that particular form of employment. Table 2.4 shows that, after grossing up the sample figures using the multipliers provided by the NSS, the SR as RSWE of men from the FC and the SC were, respectively, 32.4% and 18.2% meaning that 32.4% of men from the FC, but only 18.2% of men from the SC, were RSWE. In stark contrast, the SR as CWE of men from the FC and the SC were, respectively, 12.3% and 47.8% meaning that only one in eight men from the FC, but nearly one in two men from the SC, were CWE.

The values of inequality — defined by $J$ in equation (2.11) — calculated using the 68$^{th}$ round SR for the different groups, shown in Table 2.4, were 5.2, 9.0 and 1.2, respectively, for RSWE, CWE, and OAW. This suggests that the highest inequality was associated with CWE — with nearly half of SC and nearly one-third of ST men as CWE — and the lowest inequality was with respect to OAW with approximately four out of ten ST, OBC-NM, Muslim and FC men being OAW. Sandwiched in between these extremes was RSWE: the SR for this category was high for men from the FC but fairly evenly distributed between men from the other groups.

To put these results in perspective, in an earlier study (Borooah, 2001) I computed the values of employment inequality (J values) for Northern Ireland in the days when the Catholic share in the workforce was well short of its share in the labour force. This shortfall, in turn, generated debate about labour market discrimination which then spawned the Equal Opportunities legislation that has utterly transformed the country's labour market. These results, which are reproduced in Table 2.5, show that inequality in the Indian labour market with respect to CWE exceeds, and inequality with respect to RSWE matches, labour market inequality in Northern Ireland in the darkest days of its sectarian conflict.

Now from the definition of $e_m$, above:

$$e_m = H_m / N_m = \left(H_m / N_m\right)\left(N / H\right)\left(H / N\right) = (H_m / H)(N / N_m)(H / N) = \left(h_m / n_m\right)\overline{e} \qquad (2.10)$$

where: $h_m = H_m / H$ and $n_m = N_m / N$ are, respectively, group *m's* share of higher education attendees and of the population, substituting equation (2.10) in equation (2.9) yields:

$$J = \log(\overline{e} / \hat{e}) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log(e_m) = \log(\overline{e}) - \sum_{m=1}^{M} n_m \log\left[\frac{h_m}{n_m}\overline{e}\right] = -\sum_{m=1}^{M} n_m \log\left[\frac{h_m}{n_m}\right] \qquad (2.11)$$

From equation (2.11), inequality is minimised when *J=0*. This occurs when $n_m = h_m$, that is when each group's share in the "population" *($n_m$)* is equal to its share as RSWE ($h_m$). Inequality is at a maximum when one group has complete access (say group 1) with all access denied to the other groups ($h_1 = 1,\ h_2 = h_3 ... = h_m = 0$). Then $J_{\max} = -n_1 \log(1 / n_1) = n_1 \log(n_1)$ and, therefore, $0 \le J \le n_1 \log(n_1)$.

The inequality measure, *J*, of equation (2.11) has — along the lines suggested by Bourguignon (1979) — an appealing interpretation. If social welfare is the sum of identical and concave group utility functions whose arguments are $e_m$ then social welfare is maximised when $e_m$ — the success rate of a group — is the same for every group. If the utility functions are of the logarithmic form — that is, $U(e_m) = \log(e_m)$ — then *J* represents the distance between the *maximum* level of social welfare ( $\log(\overline{e})$ ) and the *actual* level of social welfare ( $\sum_{m=1}^{M} n_m \log(e_m)$ ): social welfare is maximised when access inequality is minimised. This theme of the link between social welfare and inequality forms the basis of much of this book and is explored in some detail in subsequent chapters.

## 2.7 Conclusions

The results reported in this chapter relate to differences between job-seekers from different social groups in terms of their risk of securing certain employment outcomes. Of particular interest is the most desirable outcome relating to jobs which pay a regular salary or wage. As Table 2.4 shows, nearly a third of FC men, compared to less than a fifth of SC men, were successful in securing such jobs. The inter-group disparities noted in this chapter are of a factual nature. No view is adduced as to the roots of these disparities. They may be due to inter-group disparities in attributes that are necessary for regular employment and the differences that we observe are, therefore, due to the "unequal treatment of unequals" or, to coin a word, due to "meritification". On this interpretation, men from the FC meet with greater success because they are better qualified for regular employment.

However, another, more insidious, reason for observing inter-group disparities in employment outcomes with respect to "good" jobs is the "unequal treatment of equals". Candidates are rejected because they belong to certain castes or religions even though they may be otherwise qualified to hold down these jobs. In such cases, disparities in employment outcomes can legitimately be regarded as being due to discrimination.

Such clear cut distinctions between meritification and discrimination are, however, rarely available in assessing actual labour market situations. Part of the reason that SC males are relatively unsuccessful, vis-à-vis their FC counterparts, in securing regular employment is due to them being less qualified: the NSS 68th round shows that, of men aged 21–55 years, only 2% of SC men compared to 8% of FC men were graduates, and only 20% of SC men compared to 44% of FC men had studied up to a secondary level. But, it would be naïve to believe that this was the only reason why men from the FC were more successful in securing regular jobs than their SC counterparts. The other part of the explanation for the relative lack of success of SC men is discrimination: other things being equal, employers prefer to engage persons from the upper castes than persons from the lower castes, and Hindus rather than Muslims. The task of the analyst is to use available data to assess the relative strength of these two aspects — meritification and discrimination — in explaining inter-group disparities in employment outcomes. This is the purpose of the next chapter.